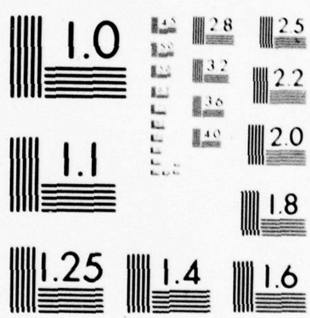
AD-A055 032 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1
A NEW CLASS OF FEASIBLE DIRECTION METHODS.(U)
SEP 77 A BEN-TAL, S ZLOBEC N00014-67-A-0126-0009

UNCLASSIFIED CCS-311 NL

1 OF 1
AD
A055 032

END
DATE
FILMED
7-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A055032

12

FOR FURTHER TRAN

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

78 06 06 057

(12)

HA008362

Research Report CCS 311

A NEW CLASS OF FEASIBLE DIRECTION METHODS

by

A. Ben-Tal*
S. Zlobec**

Extension and Revision of CCS 216

September 1977

*Department of Computer Science, Technion, Haifa, Israel.

**Department of Mathematics, McGill University, Montreal, Quebec, Canada H3C 3G1.

This research was partly supported by the National Research Council of Canada and by Project No. NR047-021, ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

ABSTRACT

This paper introduces a new class of feasible direction methods for solving differentiable convex programs with nonlinear convex constraints. Unlike many presently used methods, the ones introduced here are not based on the Fritz John or the Kuhn-Tucker theory but rather on two recent characterizations of optimality without a constraint qualification. The new methods are capable of generating feasible directions of descent along the boundary of the feasible set and they consistently give directions of steeper descent than many popular methods. This is achieved by solving only one linear program at each iteration. The new methods are particularly useful in solving large sparse convex programs; some of the programs tested had 100 variables and 50 nonlinear constraints. Moreover, the new methods are applicable whether or not Slater's condition or any other constraint qualification is satisfied.

1. INTRODUCTION

Many numerical methods for solving differentiable convex programs with nonlinear constraints are based on the necessary conditions for optimality such as the Fritz John condition (e.g. [14]). In the presence of Slater's constraint qualification the Fritz John condition is equivalent to the Kuhn-Tucker condition (e.g. [3] [14]) and it characterizes the optimality of a feasible point. It is well known that many popular feasible direction methods (e.g. [13],[23],[24],[25]) produce rather slow convergence when applied to convex programs with sparse nonlinear constraints. The main reason is that such methods generate at each iteration a direction of decrease always pointing in the interior of the feasible set and therefore a movement along the boundary of the feasible set is excluded. (See Exanmple 2 below.)

This paper introduces a new class of feasible direction methods. These methods are based on two different complete characterizations of optimality for convex programs stated in [1] and [3]. Some of the properties of the new methods are that they are capable of generating feasible directions of decrease along the boundary of the feasible region and that they consistently give a feasible direction of steeper descent than many Zoutendijk-type methods. The latter is achieved by solving only one linear program at each iteration. The fact that the new methods allow movement along the boundary of the feasible region is particularly useful in the case of programs with sparse nonlinear constraints because the feasible sets of such programs normally have "flat" parts. A movement along the flat parts can significantly speed up the convergence. Another important property of the new methods is that they can solve large sparse programs. Several of the pro-

grams tested had 100 variables and 50 nonlinear constraints. Some of the methods introduced here are also applicable to convex programs whether or not Slater's condition or any other constraint qualification is satisfied. This proves useful in the numerical treatment of programs arising in multi-criteria decision making processess because, as is well-known (e.g. [2], [17]), such programs generally lack Slater's condition.

Section 2 provides a motivation for the paper. In particular it is shown that the feasible direction methods based on the necessary conditions of optimality generally terminate at a nonoptimal point, if Slater's condition does not hold (see Example 1), and also that they point towards the interior of the feasible set, if Slater's condition does hold (see Example 2). Theoretical background for the "Method of Elimination of Linear Programs" is given in Section 3. The method at every iteration generates several direction of decrease and then, among them, it chooses one which is locally of the steepest descent. This method, studied in Section 4, is particularly useful for convex programs with sparse constraints which are strictly convex in their "actual variables". It can solve problems regardless of a constraint qualification assumption. Section 5 introduces a fundamentally different type of numerical method based on a parametric characterization of optimality given in [1]. The method is formulated for a rather large class of convex programs, namely those having "faithfully convex" constraints and in the presence of Slater's condition. An overall computational experience and two solved test programs are given in Section 6.

2. MOTIVATION

Let $f^o, f^1, \ldots, f^p$ be functions: $R^n \to R$. Consider the mathematical programming problem

$$
\text{(P)} \quad \begin{aligned} &\text{Min } f^o(x) \\ &\text{s.t.} \\ &f^k(x) \le 0, \quad k \in P \triangleq \{1,2,\ldots,p\} \end{aligned}
$$

where $x = (x_1, x_2, \ldots, x_n)^t$. It is assumed that an optimal solution x^* exists and that the functions $\{f^k : k \in \{0\} \cup P\}$ are convex and differentiable. Associated with problem (P) is the *feasible set*

$$F \triangleq \{x \in R^n : f^k(x) \le 0, \quad k \in P\}.$$

For $x \in F$ we denote by $P(x)$ the set of *active constraints*, i.e.

$$P(x) \triangleq \{k \in P : f^k(x) = 0\}.$$

In order to calculate an optimal solution x^* of (P) one can apply a *feasible direction method* (e.g. [13],[22],[23],[24],[25]). First we recall that a vector $d \in R^n$ is a *feasible direction at* $x \in F$ if

$\exists\, \bar{\alpha} > 0$ such that $x + \alpha d \in F$ for all $0 \le \alpha \le \bar{\alpha}$.

A feasible direction method iterates as follows: From a point $x^\ell \in F$, use a mapping T to determine a feasible direction d^ℓ. Then apply a mapping M to minimize the objective function f^o on a segment of the ray emanating from x^ℓ in the direction d^ℓ. Thus a new feasible point

$$x^{\ell+1} = MTx^\ell = M(x^\ell + \alpha d^\ell) = x^\ell + \alpha_\ell d^\ell$$

is obtained, where the *step-size* α_ℓ is a solution of the (one-dimensional) problem

$$(S,\ell) \qquad \text{Min}\,\{f^o(x^\ell + \alpha d^\ell): \alpha \ge 0,\ x^\ell + \alpha d^\ell \in F\}.$$

The direction d^ℓ is typically chosen to be a *direction of descent*, i.e. d^ℓ satisfies (for a nonoptimal x^ℓ)

$$(d^\ell)^t \nabla f^o(x^\ell) < 0 .$$

In Zoutendijk's classical method a feasible direction of descent at $x \in F$ is found by solving the linear program

(Z)

$$\text{Max}\ \lambda$$

s.t.

$$d^t \nabla f^k(x) + \lambda \le 0, \quad k \in \{0\} \cup P(x)$$

$$|d_i| \le 1, \quad i = 1,\ldots,n.$$

If an optimal solution of (Z) is (d^z, λ^z) and $\lambda^z > 0$, then d^z is a feasible direction of descent.

There are several variants of (Z) (see e.g. [23, p.68] and [13, p. 246]). We will call (Z) and its variants, the *Zoutendijk method* (abbreviated: Z-method). The convergence properties and problems of "jamming" (or "zig-zagging") of the Z-method have been widely discussed in the literature (e.g. [13],[22],[23],[24],[25]). Many currently used methods for solving convex programs (e.g. [13],[16],[19],[21],[25]) are based on the classical Fritz John or Kuhn-Tucker theories (e.g. [11],[14]). Whenever these theories fail to characterize optimal solutions of (P), such methods generally fail to produce an optimal solution or have a pathological behaviour. Let us first discuss briefly the situation when the Kuhn-Tucker theory fails; the case when it does not fail will be treated later. We will discuss in this paper only the Zoutendijk method. However, the basic points of the discussion apply also to the above mentioned methods.

(i) We pose the following question: Does the Z-method generally solve problem (P), with nonlinear constraints, in the absence of *Slater's condition:*

(1) $$\exists\, \hat{x} \text{ such that } f^k(\hat{x}) < 0, \quad k \in P.$$

The answer to this question is negative. If (1) is not satisfied, then at a nonoptimal feasible point x^{ℓ}, (Z) may give $\lambda^{\ell} = 0$ and the method terminates at the nonoptimal x^{ℓ}. This is illustrated by the following example.

Example 1. Consider the problem

$$\text{Min } f^{o}(x) = x_1 + x_2 + x_3$$

s.t.

$$f^1(x) = x_1^2 + x_2^2 - 2 \le 0$$

$$f^2(x) = (x_1-2)^2 + (x_2-2)^2 - 2 \le 0$$

$$f^3(x) = x_3^2 - 2x_3 \le 0$$

$$x = (x_1, x_2, x_3)^t .$$

The feasible set is

$$F = \left\{ \begin{pmatrix} 1 \\ 1 \\ x_3 \end{pmatrix} : 0 \le x_3 \le 2 \right\}$$

and the unique optimal solution is $x^* = (1,1,0)^t$.

Let the initial approximation be $x^o = (1,1,1)^t$. Problem (Z) is here

$$\text{Max } \lambda$$

s.t.

$$d_1 + d_2 + d_3 + \lambda \le 0$$

$$2d_1 + 2d_2 \qquad + \lambda \le 0$$

$$-2d_1 - 2d_2 \qquad + \lambda \le 0$$

$$|d_i| \le 1, \quad i = 1,2,3.$$

The optimal value is $\lambda^z = 0$ and an optimal solution, produced by the simplex method for linear programming, is $d^z = 0$, $\lambda^z = 0$. Thus the next approximation is

$$x^1 = x^o + \alpha d^z = x^o$$

and the algorithm terminates at x^o, a nonoptimal point.

Let us note that among infinitely many optimal solutions of problem (Z) in this example there is also an optimal solution with $d_3 < 0$, which generates d^z pointing in the right direction. However, this solution is not generally produced by the simplex method, so the Z-method generally fails.

In spite of the popular belief, the situations in which the Kuhn-Tucker theory does not characterize optimality are numerous and they appear naturally. The areas of convex programming in which Slater's condition is never satisfied and the optimal solutions are not generally the Kuhn-Tucker points include mathematical programming formulations of Pareto optimization (e.g.[2],[4]), lexicographic multicriteria decision making problems (e.g. [12]), Chebyshev solutions of inconsistent programs (e.g. [17]), programs with an ordered set of constraints (e.g. [7]) and theory of games with information exchange (e.g. [7]). Numerical methods presented in Section 4 can solve (at least in principle) such programs together with those for which a constraint qualification is satisfied.

(ii) Let us now consider the problems for which Slater's condition *is* satisfied. The optimal solutions of problem (P) are now the Kuhn-Tucker points and they are obtainable by the Z-method. However, the Z-method produces at each step only one feasible direction of decrease, always pointing to the interior of the feasible set F. This may result in a poor convergence, especially in the case of sparse nonlinear constraints of problem (P). One such situation will now be demonstrated.

<u>Example 2</u>. Consider the program

$$\text{Min } f^0(x) = x_1 + x_2$$

s.t.

$$f^1(x) = -x_1 - \frac{2}{5}x_2 + \frac{2}{5}x_2^2 \le 0$$

$$f^2(x) = x_1^2 - 1 \le 0$$

$$f^3(x) = (x_2 - 1)^2 - 1 \le 0 .$$

Definition 1. Let $f^k : R^n \to R$. Let $[k] \subset \{1,\ldots,n\}$ denote the index set of the variables $\{x_j\}$ on which f^k actually depends:

$$[k] \overset{\Delta}{=} \{j\text{: There exist } x_i = \xi_i,\ i \neq j, \text{ such that the function } f^k(\xi_1,\ldots,\xi_{j-1},.,\xi_{j+1},\ldots,\xi_n) \text{ is not a constant.}\}.$$

For any $x \in R^n$ the subvector $x_{[k]}$ is obtained by deleting the components $\{x_j: j \notin [k]\}$. The *restriction* $f^{[k]}$ of f^k is the function $f^{[k]}: R^{\mathrm{card}[k]} \to R$ obtained by restricting f^k to $x_{[k]}$.

Example 3. Consider the constraint functions of the problem from Example 1,

$$f^1(x) = x_1^2 + x_2^2 \qquad -2$$

$$f^2(x) = (x_1-2)^2 + (x_2-2)^2 - 2$$

$$f^3(x) = \qquad x_3^2 - 2x_3$$

$$x = (x_1,x_2,x_3)^t.$$

All these functions are considered, in the problem, as functions defined on R^3. However, f^1 and f^2 actually depend only on x_1 and x_2 (i.e. f^1 and f^2 are constant with respect to x_3), while f^3 actually depends only on x_3 (i.e. f^3 is constant with respect to x_1 and x_2). Therefore

$$[1] = [2] = \{1,2\} \text{ and } [3] = \{3\}.$$

Furthermore, for $k = 1,2,3$, the subvector of x consisting only of the

actual variables of f^k is denoted by $x_{[k]}$. Here

$$x_{[1]} = x_{[2]} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{and} \quad x_{[3]} = (x_3).$$

The *restriction* of f^k, $k = 1,2,3$ is the function f^k considered only as the function acting on the subspace determined by the actual variables. In this example all three functions f^k, $k = 1,2,3$ are convex, but neither is strictly convex. However, all three functions considered as functions of their actual variables, i.e. all the restrictions $f^{[k]}$, $k = 1,2,3$ *are* strictly convex. (We recall that a function $f: R^m \to R$ is *strictly convex* if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

for all $x,y \in R^m$, $x \neq y$, $0 < \lambda < 1$. Note that $f(x_1) = x_1^2$ is strictly convex when considered as $f: R \to R$, but it is *not*, when considered as $f: R^2 \to R$, i.e. when $f(x_1,x_2) = x_1^2$.)

The set of all functions $f^k: R^n \to R$ with strictly convex restrictions $f^{[k]}$ is larger (by inclusion) than the set of all strictly convex functions $f^k: R^n \to R$. The following theorem is stated for problem (P) whose constraints have strictly convex restrictions.

THEOREM 1. *Let* $\{f^k: k \in \{0\} \cup P\}$ *be differentiable convex functions:* $R^n \to R$, x^* *be a feasible solution of problem* (P) *and for all* $k \in P(x^*)$ *let the functions* $f^{[k]}$ *be strictly convex. For a given subset* Ω *of*

$P(x^*)$ let the $(n+1)$-tuple $(\lambda(\Omega), d(\Omega))$ be an optimal solution of the linear program

(L, Ω)

$$\text{Max } \lambda$$

s.t.

(2) $$d^t \nabla f^0(x^*) + \lambda \le 0$$

(3) $$d^t \nabla f^k(x^*) + \lambda \le 0, \quad k \in \Omega$$

(4) $$d_{[k]} = 0, \quad k \in \Omega^* \triangleq P(x^*) \setminus \Omega$$

(5) $$|d_i| \le 1, \quad i = 1, \ldots, n.$$

Then

(a) *x^* is an optimal solution of (P) if, and only if, for every $\Omega \subset P(x^*)$, $\lambda(\Omega) = 0$;*

(b) *a vector $\bar{d} \in R^n$ is a feasible direction of descent at x^* if, and only if, there exist a subset $\bar{\Omega}$ of $P(x^*)$ and $\bar{\lambda} > 0$ such that $(\bar{\lambda}, \bar{d})$ is a feasible solution of $(L, \bar{\Omega})$.*

Proof. It has been shown in [3, Corollary 2] that x^* is an optimal solution of problem (P), under the assumptions of Theorem 1, if, and only if, for every subset Ω of $P(x^*)$ the system

(P, Ω)

$$d^t \nabla f^0(x^*) < 0$$

$$d^t \nabla f^k(x^*) < 0, \quad k \in \Omega$$

$$d_{[k]} = 0, \quad k \in \Omega^*$$

is inconsistent. (The proof of this result is long and therefore it is not repeated here.) But the inconsistency of (P,Ω) is equivalent to $\lambda^* = 0$ being the optimal value of the linear programming program

$$\begin{aligned}
&\text{Max } \lambda \\
&\text{s.t.} \\
&\quad d^t \nabla f^o(x^*) + \lambda \le 0 \\
&\quad d^t \nabla f^k(x^*) + \lambda \le 0, \quad k \in \Omega \\
&\quad d_{[k]} = 0, \quad k \in \Omega^*.
\end{aligned}$$

The normalization condition (5) can be added to the above program without changing its optimal value. This proves the statement (a). Assume now that there exists $\bar{\Omega} \subset \mathcal{P}(x^*)$ and $\bar{\lambda} > 0$ such that $(\bar{\lambda},\bar{d})$ is a feasible solution of $(L,\bar{\Omega})$. Then $(P,\bar{\Omega})$ holds for $d = \bar{d}$, implying that $\bar{d}$ is a feasible direction of descent at x^*. Conversely, if $\bar{d}$ is a feasible direction of descent at x^*, then the system

$$\begin{cases}
\bar{d}^t \nabla f^o(x^*) < 0 \\
\bar{d}^t \nabla f^k(x^*) \le 0, \quad k \in \mathcal{P}(x^*) \\
\text{with equality only for those } k \in \mathcal{P}(x^*) \text{ for} \\
\text{which } \bar{d}_{[k]} = 0
\end{cases}$$

is consistent. This implies the existence of a subset $\bar{\Omega}$ of $\mathcal{P}(x^*)$ such that $(P,\bar{\Omega})$ is consistent and further the existence of $(\bar{\lambda},\bar{d})$, with $\bar{\lambda} > 0$, which is a feasible solution of $(L,\bar{\Omega})$. The statement (b) is proved.

□

Theorem 1 is important for the following reasons:

(i) It gives a *characterization* of optimality without any constraint qualification assumption. In particular, using Theorem 1, one can treat the lexicographic and Pareto optimization problems and establish whether a feasible solution x^* is optimal.

(ii) It explains why the Z-method generally terminates at a nonoptimal point in the absence of Slater's condition. If a feasible point x^* is not optimal, then, by the part (a) of Theorem 1, there *exists* a subset $\bar{\Omega}$ of $P(x^*)$ such that the optimal value of the corresponding linear program $(L,\bar{\Omega})$ is positive. This $\bar{\Omega}$ can be *any* subset of $P(x^*)$, not necessarily $\bar{\Omega} = P(x^*)$, unless Slater's condition is satisfied.

(iii) It explains why the Z-method, in some situations, produces a slow convergence even in the presence of Slater's condition. The reason is in the following: If x^* is a nonoptimal feasible point, then the optimal value of the linear program $(L,P(x^*))$ is positive and one obtains a feasible direction of decrease d^z. However, there are, generally, still other subsets Ω of $P(x^*)$ for which the optimal solution of (L,Ω) is positive and which also generate (different) feasible directions of decrease. In general, there are many such directions (at most $2^{\text{card } P(x^*)}$) and d^z, being only one of many, is not generally the best. For instance, suppose that at a feasible point x^*, two subsets, say Ω_1 and $\Omega_2 = P(x^*)$, have the property that the optimal values λ_1 and λ_2 of (L,Ω_1) and (L,Ω_2), respectively, are positive, in which case the two directions

d^1 and $d^2 = d^z$ are obtained. Now, if one chooses for the feasible direction of descent d^1, if $\lambda_1 > \lambda_2$, or d^2, if $\lambda_1 \le \lambda_2$, he gets locally a better feasible direction of decrease than, or at least as good as, by applying the Z-method (which produces only one direction $d^2 = d^z$). One expects to obtain even better feasible direction of decrease if three, rather than two, subsets are considered at x^*. In general, the higher is the number of subsets Ω considered, the better is the direction of descent. If all subsets are considered, one gets locally the best feasible direction of descent, by the statement (b) of Theorem 1.

In view of the above remarks, it is clear that one can use Theorem 1 to formulate a new class of feasible direction methods. These new feasible direction methods solve problem (P) without any reference to the Kuhn-Tucker theory. If, at each iteration x^ℓ, one considers k linear programming problems (L, Ω_1), (L, Ω_2), ..., (L, Ω_k), $\Omega_i \subset P(x^\ell)$, $i = 1,2,\ldots,k$, which have the optimal values $\lambda_1,\lambda_2,\ldots,\lambda_k$ all positive, and if the feasible direction of decrease is chosen to be the one generated by the linear program having the largest optimal value, then we say that the feasible direction method is of *order* k. The subsets Ω will be considered in the decreasing order, i.e. the one with the largest cardinality is considered first, then one with the second largest cardinality is considered, etc. Note that, whenever Slater's condition is satisfied, the feasible direction of order $k = 1$ is exactly the Z-method. (If Slater's condition is satisfied, $\Omega = P(x^\ell)$ is the largest subset of $P(x^\ell)$ and $\lambda(P(x^\ell)) = \lambda^z > 0$.)

Let us revisit Examples 1 and 2. Theorem 1 will now be applied to the two problems in order to illustrate the points (ii) and (iii). New feasible direction methods, which are based on Theorem 1, will be formally introduced and studied in Section 4.

Example 4. The Z-method has terminated in Example 1 at the nonoptimal point $x^o = (1,1,1)^t$. This has happened because Slater's condition is not satisfied and the optimal value of $(L, P(x^o))$ is zero. However, by Theorem 1, we know that there exists a subset $\bar{\Omega}$ of $P(x^o)$ such that the optimal value of $(L, \bar{\Omega})$ *is* positive. Indeed, for $\bar{\Omega} = \phi$ (the empty set),

(L,ϕ)

$$\begin{aligned} &\max \lambda \\ \text{s.t.}\quad & d_1 + d_2 + d_3 + \lambda \le 0 \\ & d_1 = d_2 = 0 \\ & -1 \le d_i \le 1,\ i = 1,2,3 \end{aligned}$$

has the unique optimal solution $\lambda = 1,\ d_1 = d_2 = 0,\ d_3 = -1$. The corresponding direction $d = (0,0\text{-}1)^t$ is feasible and it points towards the optimal solution $x^* = (1,1,0)^t$.

Example 5. Consider again the program introduced in Example 2. Starting from $x^0 = (\frac{1}{2},0)^t$ and specifying $\Omega = \emptyset$, the program $(L,\emptyset)$ gives the feasible direction $d^0 = (-1,0)^t$, pointing towards the optimal solution $x^* = 0$. Thus the program is solved in only one iteration.

If Slater's condition holds for problem (P) then the family of linear programs $\{(L,\Omega): \Omega \subset P(x^*)\}$ in Theorem 1 can be replaced by the single program $(L,P(x^*))$. This is a well-known result, which is stated below for the sake of completeness.

<u>Proposition 1</u>. *Let* $\{f^k: k \in \{0\} \cup P\}$ *be differentiable convex functions:* $R^n \to R$ *and* x^* *be a feasible solution of problem* (P). *If Slater's condition holds for problem* (P), *then* x^* *is an optimal solution of* (P) *if, and only if, the optimal value of the linear program*

$(L,P(x^*))$

$$\begin{aligned} &\text{Max } \lambda \\ \text{s.t.}\quad & \\ &d^t\nabla f^o(x^*) + \lambda \le 0 \\ &d^t\nabla f^k(x^*) + \lambda \le 0,\ k \in P(x^*) \\ &|d_i| \le 1,\ i = 1,\dots,n \end{aligned}$$

is zero.

In order to check whether Slater's condition holds one can use the following proposition.

<u>Proposition 2</u>. *Slater's condition holds for problem* (P) *if, and only if, for an arbitrary fixed feasible point* $x \in F$, *the optimal value* λ^* *of the linear program*

(CQ)

$$\begin{aligned} &\text{Max } \lambda \\ \text{s.t.}\quad & \\ &d^t\nabla f^k(x) + \lambda \le 0,\ k \in P(x) \\ &|d_i| \le 1,\ i = 1,\dots,n \end{aligned}$$

is positive.

4. THE METHOD OF ELIMINATION OF LINEAR PROGRAMS

We will first study the method in which a direction of decrease is determined by solving only one linear program. This method is termed the Method of Elimination of Linear Programs of First Order (abbreviated: MELP 1).

If Slater's condition is satisfied, then MELP 1 coincides with the Z-method. If Slater's condition does not hold for problem (P), then $\lambda^* = 0$ is the optimal value of (CQ), by Proposition 2. This implies that $\lambda^* = 0$ is also the optimal value of problem (Z). But (Z) is exactly $(L, P(x))$. This means that, if Slater's condition does not hold, $\lambda(P(x)) = 0$ for every $x \in F$. Therefore, in order to test optimality of $x^* \in F$ using Theorem 1, only proper subsets Ω of $P(x^*)$ need checking. This calls for solving $2^{\mathrm{card}\, P(x^*)} - 1$ linear programs (at most). However, in many cases one can establish the zero optimal value of the programs (L, Ω) without actually solving these programs. In order to demonstrate this assertion, recall that in case of programs with strictly convex restrictions the program (L, Ω) has the zero optimal value if, and only if, the system (P, Ω) is inconsistent. Whenever the components of d which correspond to the nonzero components of $\nabla f^0(x^*)$ or $\nabla f^k(x^*)$, for at least one index $k \in \Omega$, are annihilated by the requirement $d_{[k]} = 0$, $k \in \Omega^*$, then the system (P, Ω) contains the contradictory statement $0 < 0$, i.e. the system (P, Ω) is inconsistent and the program (L, Ω) has the optimal value zero. If this happens for a particular subset Ω of $P(x^*)$, we say that the program (L, Ω) is *eliminated*. The above condition can be formulated as follows:

<u>Proposition 3.</u> *If for a given proper subset* Ω *of* $P(x^*)$ *the elimination condition:*

$$(6)\quad \begin{cases} \textit{There exists } k \in \{0\} \cup \Omega \textit{ such that} \\ \left\{ i : \frac{\partial}{\partial x_i} f^k(x^*) \neq 0 \right\} \subset \bigcup_{j \in \Omega^*} [j] \end{cases}$$

is satisfied, then $\lambda(\Omega) = 0$. *In fact, if (6) is satisfied for* $k = k_o$, *then* $\lambda(\bar{\Omega}) = 0$ *for every nonempty subset* $\bar{\Omega}$ *of* $P(x^*)$ *such that*

$$(7)\qquad k_o \in (\{0\} \cup \bar{\Omega}) \subset (\{0\} \cup \Omega).$$

Proof. **First note that the constraint $\lambda \geq 0$ may be added to the constraints of (L, Ω) because $\lambda = 0$, $d = 0$ is a feasible solution. If (6) holds for $k = k_o$, then the elimination condition and the conditions (7) imply that**

$$d^t \nabla f^{k_o}(x^*) = 0$$

and hence (3), or (2), becomes $\lambda \leq 0$. Therefore the optimal value of (L, Ω) is clearly $\lambda(\Omega) = 0$. Note that

$$(\{0\} \cup \bar{\Omega}) \subset (\{0\} \cup \Omega) \Rightarrow \bar{\Omega} \subset \Omega .$$

But

$$\bar{\Omega} \subset \Omega \Rightarrow \bar{\Omega}^* \supset \Omega^* \Rightarrow \bigcup_{j \in \bar{\Omega}^*} [j] \supset \bigcup_{j \in \Omega^*} [j]$$

and hence the elimination condition holds for $\bar{\Omega}$. Therefore $\lambda(\bar{\Omega}) = 0$. $\square$

In general, the "fuller" is the problem (i.e. the more actual variables appear in the objective function and constraints of (P)) the more programs (L, Ω) are eliminated. If the elimination condition holds for *all* subsets of $P(x^*)$, then x^* is optimal, by the part (a) of Theorem 1. Let us also note that the idea of elimination of linear programs is meaningful for problem (P) even if Slater's condition does hold. The elimination condition (6) will now be illustrated by an example.

<u>Example 6</u>. Consider

$$\min f^o(x) = x_2 + e^{-x_4}$$

s.t.

$$f^1(x) = e^{x_1} + e^{x_2} + e^{-x_3} + e^{x_4} - 4 \le 0$$

$$f^2(x) = e^{x_1} + (1-x_3)^2 + e^{-x_4} - 3 \le 0$$

$$f^3(x) = x_1^2 + e^{-x_2} + (1+x_4)^2 - 2 \le 0$$

$$f^4(x) = x_1^2 + e^{x_2} + e^{x_3} + e^{-x_4} - 3 \le 0$$

$$f^5(x) = e^{-x_1} + e^{x_2} + 2e^{x_3} - 5 \le 0 .$$

Let $x^* = 0$. Then $P(x^*) = \{1,2,3,4\}$. The number of proper subsets $\{\Omega \subset P(x^*)\}$ is $2^4 - 1 = 15$. Consider, for instance, $\Omega = \{1\}$. Then $\Omega^* = \{2,3,4\}$ and the system

(P,{1})

$$d_2 - d_4 < 0 \qquad (d^t \nabla f^o(x^*) < 0)$$

$$d_1 + d_2 - d_3 + d_4 < 0 \qquad (d^t \nabla f^1(x^*) < 0)$$

$$d_1 = d_3 = d_4 = 0 \qquad (d_{[2]} = 0)$$

$$d_2 = d_4 = 0 \qquad (d_{[3]} = 0)$$

$$d_2 = d_3 = d_4 = 0 \qquad (d_{[4]} = 0)\ .$$

is clearly inconsistent. The program (L,{1}) is thus eliminated. Instead of writing down all the systems $\{(P, \Omega): \Omega \subset P(x^*)\}$ to be checked for elimination, one can associate with the above problem its *incidence matrix* $A = (a_{kj})$, the components of which are

$$a_{kj} = \begin{cases} 0 & \text{if } \dfrac{\partial}{\partial x_j} f^k(x^*) = 0 \\ 1 & \text{otherwise}\ . \end{cases}$$

Here

$$\frac{\partial}{\partial x_j} f^k(0)$$

	k \ j	1	2	3	4	
	0	0	1	0	1	← k = 1 (Ω = {1})
	1	1	1	1	1	
	2	1	0	1	1	
f^k	3	0	1	0	1	
	4	0	1	1	1	
	5	1	1	1	0	← nonbinding constraint

For $\Omega = \{1\}$, one considers the above table and concludes that the elimination condition (6) is satisfied for $k = 0$ (also for $k = 1$). Therefore, program $(L,\{1\})$ is eliminated. Checking of all the subsets Ω of $P(x^*)$ in this manner reveals that all the programs (L,Ω) are eliminated, except three, which are $(L,P(x^*))$, (L,ϕ) and $(L,\{1,3,4\})$. Thus only programs $(L,P(x^*))$, (L,ϕ) and $(L,\{1,3,4\})$ have to be actually solved in order to check whether $x^* = 0$ is optimal.

Let us note that the last part of Proposition 3 is quite useful in the process of elimination. For instance, as soon as one has established that the program $(L,\{1,2,3\})$ is eliminated, one can conclude that the programs $(L,\{1,2\})$, $(L,\{1,3\})$, $(L,\{2,3\})$, $(L,\{1\})$, $(L,\{2\})$ and $(L,\{3\})$ (but not (L,ϕ)!) are eliminated as well. (Specify $\Omega = \{1,2,3\}$, $\Omega^* = \{4\}$ and $k_o = 0$ in Proposition 3. Relation (7) becomes $0 \in (\{0\} \cup \bar{\Omega}) \subset \{0,1,2,3\}$ which is satisfied for nonempty $\bar{\Omega}$ equal to $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1\}$, $\{2\}$ or $\{3\}$.) This suggests that in the process of elimination one should start with subsets of cardinality $q = P(x^*)$ and proceed to subsets of smaller cardinality.

If Slater's condition does not hold then the MELP 1 iterates as follows, given a current point $x^\ell \in F$:

(a) Let $\Omega_o = P(x^\ell)$.

(b) Check the elimination condition (6) on all subsets Ω of $P(x^{\ell})$, for which it is not established that $\lambda(\Omega) = 0$, $\Omega \neq \Omega_o$, and for which card $\Omega \leq$ card Ω_o, starting with a subset Ω of maximal cardinality. If an Ω is not eliminated, go to (c). If all Ω's are eliminated, x^{ℓ} is optimal.

(c) Set $\Omega_o = \Omega$ and solve (L, Ω_o). If $\lambda(\Omega_o) > 0$, use $d^{\ell} = d(\Omega_o)$ as a direction of descent. If $\lambda(\Omega_o) = 0$ and $\Omega_o \neq \emptyset$, go to (b). If $\lambda(\Omega_o) = 0$ and $\Omega_o = \emptyset$, x^{ℓ} is optimal.

The MELP of Second Order (abbreviated MELP 2) determines at every point x^{ℓ} two linear programs, say (L,Ω_1) and (L,Ω_2) with nonzero solutions. If (λ_1,d^1) and (λ_2,d^2) are the nonzero solutions of these programs, then the feasible direction of decrease d^{ℓ} is chosen as follows:

$$d^{\ell} = \begin{cases} d^1 & \text{if } \lambda_1 \geq \lambda_2 \\ d^2 & \text{otherwise} . \end{cases}$$

In view of Proposition 3, it is suggested that linear programs (L, Ω) with the highest cardinality of Ω be first considered. Once d^{ℓ} is determined, one solves the one-dimensional step-size problem (S,ℓ) and obtains a new point $x^{\ell+1}$.

Remarks.

(i) Let us note that d^1 and d^2 depend on Ω, i.e. d^1 is generated by a subset Ω_1, while d^2 is generated by another subset Ω_2. At every iteration x^{ℓ}, the subsets Ω_1 and Ω_2 are generally

different. However, if Slater's condition is satisfied, one of these two subsets is always $\Omega = P(x^\ell)$. (This is a consequence of Proposition 1: if x^ℓ is not optimal, then $(L, P(x^\ell))$ has an optimal solution (λ^z, d^z) with $\lambda^z > 0$ and $d^z \neq 0$. Since $\Omega = P(x^\ell)$ has the highest cardinality of all the subsets of $P(x^\ell)$, $(L, P(x^\ell))$ is solved and d^z is picked as one of the two candidates for d^ℓ.) Thus, if Slater's condition holds, the feasible direction of decrease d^ℓ is chosen as follows:

$$d = \begin{cases} d^z & \text{if } \lambda^z \geq \lambda_2 \\ d^2 & \text{otherwise.} \end{cases}$$

Here (λ_2, d^2) is an optimal solution of (L, Ω) with Ω having the highest cardinality among all *proper* subsets of $P(x^\ell)$ and such that $d^2(\Omega) \neq 0$ in $(\lambda(\Omega), d^2(\Omega))$.

(ii) If, at $x^\ell \in F$, all (L, Ω)'s are eliminated or have optimal solutions $(\lambda(\Omega), d(\Omega))$ with $d(\Omega) = 0$, except one, then MELP 2 coincides with MELP 1 at this particular point.

The MELP of Third Order (abbreviated MELP 3) chooses at every $x^\ell \in F$ the feasible direction of decrease d^ℓ, by comparing three nonzero solutions (λ_1, d^1), (λ_2, d^2) and (λ_3, d^3). The one with maximal λ determines d^ℓ. For instance, if $\lambda_2 = \max\{\lambda_1, \lambda_2, \lambda_3\}$, then $d^\ell = d^2$.

It is obvious that one can define the MELP of a higher order, as well. The "best" feasible direction of decrease is the one chosen after solving and comparing all (noneliminated) programs (L,Ω) with nonzero optimal solutions. In Section 5 it will be shown how th "best" feasible direction of decrease can be determined at $x^{\ell} \varepsilon F$ by a different. more practical approach.

The MELP (in particular MELP 2) has produced good numerical results for sparse convex programs with constraints which have strictly convex restrictions. In case of the general convex program the MELP is much more involved and its computer implementation is not yet finalized. Nevertheless, for the sake of completness, and as a suggestion for possible future research, the MELP for general convex programs will now be briefly described.

First we recall the following concept (e.g. [3]) .

<u>Definition 2</u>. Let $f: R^n \to R$ and $x^* \epsilon R^n$. Then

$$D_f(x^*) \triangleq \{d \epsilon R^n : \exists\, \bar{\alpha} > 0 \ni f(x^* + \alpha d) = f(x^*), \forall \alpha \epsilon [0,\bar{\alpha}]\}$$

is the *cone of directions of constancy* of f at x^*.

If f is a convex differentiable function then $D_f(x^*)$ is a convex (but not necessarily closed) cone. In general, the cone of directions of constancy can be quite complicated. However, if the function f is strictly convex, strictly convex in its actual variables or linear, this cone is quite simple, as shown by the following example.

Example 7. If $f: R^n \to R$ is strictly convex, then at every $x^* \in R^n$,

$$D_f(x^*) = \{0\}.$$

If $f: R^n \to R$ is strictly convex in its actual variables $x_1,\ldots,x_k$ $(k \le n)$, then at every $x^* \in R^n$,

$$D_f(x^*) = \left\{ \begin{pmatrix} 0 \\ \vdots \\ 0 \\ d_{k+1} \\ \vdots \\ d_n \end{pmatrix} : d_i,\ i = k+1,\ldots,n \text{ arbitrary} \right\}.$$

For instance, consider $f^1: R^3 \to R$ and $f^2: R^3 \to R$ defined by

$$f^1 = x_1^2 + x_2^2 - 2, \quad f^2 = e^{-x_3} - 1.$$

Here

$$D_{f^1}(x^*) = \left\{ \begin{pmatrix} 0 \\ 0 \\ d_3 \end{pmatrix} : d_3 \text{ arbitrary} \right\}$$

and

$$D_{f^2}(x^*) = \left\{ \begin{pmatrix} d_1 \\ d_2 \\ 0 \end{pmatrix} : d_1, d_2 \text{ arbitrary} \right\}.$$

If $f: R^n \to R$ is a linear form $f(x) = a^t x + \alpha$, where $a \in R^n$ and $\alpha \in R$, then at every $x^* \in R^n$,

$$D_f(x^*)=N(a^t),$$

the null-space of a^t. More general, a function $f\colon R^n \to R$ of the form

$$f(x) = \Phi(Ax+b) + x^t a + \beta$$

where Φ is a strictly convex function: $R^m \to R$, A is an $m\times n$ matrix, $b \in R^m$, $a \in R^n$ and $\beta \in R$ has

$$D_f(x^*) = \{d \in R^n\colon\ Ad = 0,\ d^t a = 0\}\ .$$

These functions are called "faithfully convex" by Rockafellar. A class of such functions is rather large and it includes linear, strictly convex, quadratic and in fact all analytic convex functions.

The MELP is based on the following characterization of optimality stated for general convex program (P).

THEOREM 2. *Let* $\{f^k\colon\ k \in \{0\} \cup P\}$ *be differentiable convex functions:* $R^n \to R$ *and* x^* *be a feasible solution of problem* (P). *For a given subset* Ω *of* $P(x^*)$ *let the* $(n+1)$*-tuple* $(\lambda(\Omega), d(\Omega))$ *be an optimal solution of the linear program over a cone*

(C,Ω)

$$\max \lambda$$

s.t.

$$d^t \nabla f^o(x^*) + \lambda \le 0$$

$$d^t \nabla f^k(x^*) + \lambda \le 0,\quad k \in \Omega$$

$$d \in D_{f^k}(x^*),\quad k \in \Omega^*\ ;\quad |d_i| \le 1,\quad i = 1,\dots,n.$$

Then

(a) x^* *is an optimal solution of* (P) *if, and only if, for every* $\Omega \subset P(x^*)$, $\lambda(\Omega) = 0$;

(b) *a vector* $\bar{d} \in R^n$ *is a feasible direction of descent at* x^* *if, and only if, there exists a subset* $\bar{\Omega}$ *of* $P(x^*)$ *and* $\bar{\lambda} > 0$ *such that* $(\bar{\lambda}, \bar{d})$ *is a feasible solution of* (C,Ω).

Proof. Point $x^* \in F$ is an optimal solution of problem (P) if, and only if, for every subset Ω of $P(x^*)$ the system

$$d^t \nabla f^o(x^*) < 0$$

$$d^t \nabla f^k(x^*) < 0, \quad k \in \Omega$$

$$d \in D_{f^k}(x^*), \quad k \in \Omega^*$$

is inconsistent. (Proof of this statement can be found in [3, Theorem 1].) Theorem 2 is now proved in the same way as Theorem 1, the only difference being that "$d_{[k]} = 0$" is replaced by "$d \in D_{f^k}(x^*)$".

□

Note that Theorem 1 is a special case of Theorem 2, since in the case of constraint functions f^k, $k \in P$, which are strictly convex in their actual variables,

$$d \in D_{f^k}(x^*) \Leftrightarrow d_{[k]} = 0.$$

The MELP method (of order k) is formulated in the same way as before. The only difference is that instead of solving (or eliminating) linear programs (L,Ω) one has to solve (or eliminate) linear programs over cones (C,Ω) .

The main difficulty in implementing the MELP in the general case is that the cones $D_{f^k}(x^*)$, $k \in \Omega^*$ have to be calculated at every iteration. However, in many situations, as shown by Example 7, these cones are readily available.

5. THE PARAMETRIC FEASIBLE DIRECTION METHOD

The method presented in this section is based on the parametric characterization of optimality given in [1, Theorem 3] . Its main features are that it is also capable of generating directions along the boundary of the feasible region and it consistently gives a feasible direction of steeper descent than the direction generated by the Z-method. This is achieved by solving only one linear program at each iteration. Numerical experience indicates that, in spite of the additional work per iteration, the parametric feasible direction method (abbreviated PFDM) is overall superior to the Z-method (especially for large problems) since the number of iterations, required to achieve a desired accuracy, is much smaller. The method is also capable of solving large sparse convex programs.

The PFDM is designed to solve convex differentiable programs, with faithfully convex constraints (see Example 7), which satisfy Slater's condition:

$$\text{Min} \quad f^o(x)$$

s.t.

$$f^k(x) \triangleq h^k(A_k x + b^k) + x^T a^k + \alpha_k \le 0, \quad k \in P$$

(FC)

$$x^T c^\ell + \beta_\ell \le 0, \quad \ell \in L$$

$$L \le x \le U$$

where

$$h^k : R^{m_k} \to R \text{ is strictly convex}$$

$$A_k \in R^{m_k \times n}, \; b^k \in R^{m_k}, \; a^k \in R^n, \; \alpha_k \in R \; (k \in P), \; c^\ell \in R^n,$$

$$\beta_\ell \in R, \; \ell \in L, \; L \in R^n, \; U \in R^n.$$

The method is based on the following result.

<u>Theorem 3.</u> Let x be a feasible nonoptimal solution of the convex program (FC). Then there exists a sufficiently small positive scalar θ such that $d = d(\theta)$, the optimal solution of the linear program

$$\text{Min} \quad d^t \nabla f^o(x)$$

s.t.

$$d^t \nabla f^k(x) + \theta \left(|d^t a^k| + \sum_{i=1}^{p} |A_k^i d| \right) \le 0, \quad k \in P(x)$$

$$|d_i| \le 1, \quad i \in P$$

is a feasible direction of descent at x. Moreover, $0 < \theta_1 \le \theta$ implies

$$(d(\theta_i))^t \nabla f^o(x) \le (d(\theta))^t \nabla f^o(x)$$

i.e. the smaller is θ, the steeper is descent. Here A_k^i denotes the i-th row of A_k.

Proof. The result follows Theorem 3 in [1] after specifying φ^k, $k \in P$ to be the ℓ_1 norm.

□

For the use of anti-jamming procedure we introduce the ε-active index sets:

$$P_\varepsilon(x) = \{k \in P : -\varepsilon \le f^k(x) \le 0\}$$

$$L_\varepsilon(x) = \{\ell \in L : -\varepsilon \le x^T c^\ell + \beta_\ell \le 0\}$$

$$J_\varepsilon^+(x) = \{j : U_j - \varepsilon \le x_j \le U_j\}$$

$$J_\varepsilon^-(x) = \{j : L_j \le x_j \le L_j + \varepsilon\}$$

Let x be a feasible solution of (FC). The Z-method with the anti-jamming ε-active procedure for problem (FC) uses the following direction generating linear program

(Z_ε)

Max λ

s.t.

(8) $$\nabla f^k(x)d + \lambda \le 0, \quad k \in \{0\} \cup P_\varepsilon(x)$$

$$d^T c^\ell \le 0, \quad \ell \in L_\varepsilon(x)$$

$$d_j \le 0, \quad j \in J_\varepsilon^+(x)$$

$$(9) \qquad -1 \le d_j \le 1, \quad j = 1,\dots,n$$

$$\sum_{j=1}^{n} |d_j| \le \sqrt{2}$$

$$\lambda \ge 0.$$

For a given feasible direction d, the step-size generating problem for (FC) is

$$(S,\ell) \qquad \text{Min} \quad f^o(x + \alpha d)$$

$$\text{s.t.}$$

$$f^k(x + \alpha d) \le 0, \quad k \in P$$

$$\alpha_1 \le \alpha \le \alpha_2$$

$$\alpha \ge 0$$

where

$$\alpha_1 = \max\left\{ \max_{\ell \in L^-} \frac{-\beta_\ell - x^t c^\ell}{d^t c^\ell}, \ \max_{j \in I^-} \frac{U_j - x_j}{d_j}, \ \max_{j \in I^+} \frac{L_j - x_j}{d_j} \right\}$$

$$\alpha_2 = \min\left\{ \min_{\ell \in L^+} \frac{-\beta_\ell - x^t c^\ell}{d^t c^\ell}, \ \min_{j \in I^+} \frac{U_j - x_j}{d_j}, \ \min_{j \in I^-} \frac{L_j - x_j}{d_j} \right\}.$$

while

$$I^+ = \{j: \ d_j > 0\}, \quad I^- = \{j: \ d_j < 0\}$$

$$L^+ = \{\ell: \ d^t c^\ell > 0\}, \quad L^- = \{\ell: \ d^t c^\ell < 0\} \ .$$

Here we use the convention that the maximum over the empty set is $-\infty$ and the minimum over the empty set is $+\infty$. The bounds α_1 and α_2 are obtained by requiring the point $x + \alpha d$ to be feasible for the linear constraints of (FC).

Given a solution d^z of (Z_ε) define the numbers

(10) $$\theta_k = - \frac{(d^z)^t \nabla f^k(x)}{|(d^z)^t a^k| + \sum_{i=1}^{m_k} |A_k^i d^z|}, \quad k \in P_\varepsilon(x)$$

and the linear program

(L_ε)

$$\text{Min} \quad d^t \nabla f^o(x)$$

s.t.

(11) $$d^t \nabla f^k(x) + \theta_k \left(|(d^z)^t a^k| + \sum_{i=1}^{m_k} |A_k^i d^z| \right) \le 0, \quad k \in P_\varepsilon(x)$$

and all constraints (9).

Proposition 4. Let x be a feasible but nonoptimal solution of (FC), $\lambda^z > 0$, d^z an optimal solution of (Z_ε) and d^L an optimal solution of (L_ε). Then d^L is a feasible direction of steeper descent than d^z, i.e.

(12) $$(d^L)^t \nabla f^o(x) \le (d^z)^t \nabla f^o(x).$$

Proof. Since $\lambda^z > 0$

(13) $$(d^z)^t \nabla f^k(x) < 0, \quad k \in P_\varepsilon(x)$$

by (8). Further, by the assumption that f^k, $k \in P$ are faithfully convex,

$$(d^z)^t \nabla f^k(x) = [\nabla h^k(A_k x + b^k)] (A_k d^z) + (d^z)^t a^k$$

which together with (13) shows that the denominator in θ_k is nonzero. Thus θ_k is well defined and positive for every $k \in P_\varepsilon(x)$. Hence, by

Theorem 3 (see also discussion in [1]), d^L is a feasible direction. Now the definition of θ_k and the fact that d^z solves (Z_ε) imply that $d = d^z$ satisfies (9) and (11), i.e. $d = d^z$ is feasible for (L_ε) and hence (12) follows.

□

The PFDM will now be formulated for solving the program (FC).

Initialization. Specify ε_o ("ε-activity parameter") and $\varepsilon' < \varepsilon_o$ ("stopping-rule parameter"). Find an initial feasible solution x^o. Set $k = 0$.

Step 1. Set $x = x^k$, $\varepsilon = \varepsilon_k$. Solve (Z_ε). Let λ^z, d^z denote the solution. Set $\lambda = \lambda^z$, $d = d^z$.

Step 2. If $\mathcal{P}_\varepsilon(x) = \emptyset$, go to Step 4. Otherwise continue.

Step 3. Calculate θ_k by formula (10), solve (L_ε). Denote the solution by d and set $\lambda = d^t\nabla f^o(x)$.

Step 4. If $\lambda > \varepsilon'$, solve (S,ℓ). Let α^* be the optimal solution. Set $x^{k+1} = x + \alpha^* d$ and continue. Otherwise, set $x^{k+1} = x$ and go to Step 6.

Step 5. If $\varepsilon \le \varepsilon'$ set $\varepsilon_{k+1} = \frac{1}{2}\varepsilon_{k-1}$. Otherwise set $\varepsilon_{k+1} = \min(\varepsilon, \lambda)$. Go to Step 1.

Step 6. If $\varepsilon \le \varepsilon'$ stop, x^{k+1} is optimal. Otherwise set $\varepsilon_{k+1} = \min(\varepsilon, \lambda)$ and go to Step 1.

Remarks

(i) The absolute value variables in (Z_ε) are treated by the transformation

$$d_i = d_i^+ - d_i^- \ , \ |d_i| = d_i^+ + d_i^- \ , \ d_i^+ \ge 0, \ d_i^- \ge 0$$

$$d_i^+ d_i^- = 0 \ . \tag{14}$$

The simplex algorithm applied to (Z_ε) has to be modified so that the orthogonality condition (14) is satisfied. (Such modification is standard in e.g. quadratic programming algorithms.)

(ii) Introducing the transformation

$$w_i^k = A_k^i d \quad , \quad i = 1,\ldots,m_k$$

$$w_{m_k+1}^k = d^t a^k$$

the constraint (11) becomes

$$d^t \nabla f^k(x) + \theta_k \left(\sum_{k=1}^{m_k+1} |w_i^k| \right) \le 0$$

$$A_k^i d - w_i^k = 0 \qquad , \ i=1,\ldots,m_k \tag{15}$$

$$d^t a^k - w_{m_k+1}^k = 0$$

The absolute values in this system are treated as in Remark (i).

(iii) The constraints (15), and also the absolute values transformation, increase the size of the linear programming problem equivalent to (L_ε) . Nevertheless the corresponding matrix is sparse and has a special structure which should be exploited to speed up calculations.

(iv) The normalization condition $-1 \le d_j \le 1,\ j=1,\dots,n$ is computationally convenient but may cause overall slow convergence. The same can be said about the condition $\sum_{j=1}^{n} |d_j| \le 1$. The combination

$$-1 \le d_j \le 1 \ ,\ j=1,\dots,n$$

$$\sum_{j=1}^{n} |d_j| \le \sqrt{2}$$

used in the algorithm can be still described by linear constraints and it has been found satisfactory in test problems. The above normalization condition is an approximation of the normalization $\sum_{j=1}^{n} d_i^2 \le 1$.

(v) In the practical implementation of the PFDM it has been found that $\varepsilon_o \approx 0.1$ is a satisfactory choice.

(vi) Numerical experience indicates that, particularly in the early stages of the algorithm, it is preferable to solve the step-size problem not too accurately.

6. OVERALL COMPUTATIONAL EXPERIENCE

The authors have solved by the MELP and PFDM, more than one hundred convex programs with faithfully convex nonlinear constraints. The

size of the program has ranged from small to the ones with 100 variables and 50 nonlinear constraints. The overall experience suggests that the MELP 2 gives very good results, particularly for sparse programs with constraints functions having strictly convex constraints. The method also seems to be rather uneffected by jamming. The PFDM produces excellent results for any kind of faithfully convex constraints.

Let us finally demonstrate the methods by solving two nontrivial programs.

Example 8. Consider

$$\text{Min} \quad f^o(x) = x_1 - x_2 + (x_3-1)^2 + (x_4-2)^2 + (x_5-2)^2$$

s.t.

$$f^1(x) = e^{x_1} + x_2^2 - 1 \le 0$$

$$f^2(x) = x_1^2 + x_2^2 + e^{-x_3} - 1 \le 0$$

$$f^3(x) = x_1 + x_4^2 + x_5^2 - 1 \le 0$$

$$f^4(x) = x_2^2 - 2x_2 \le 0$$

$$f^5(x) = (x_1-1)^2 + x_2^2 - 1 \le 0$$

$$f^6(x) = x_1 + e^{-x_4} - 1 \le 0$$

$$f^7(x) = x_2 + e^{-x_5} - 1 \le 0.$$

One can show, that at the feasible point $x^o = 0$, the optimal value λ^* of the linear program (CQ) is zero. Therefore, Slater's condition is not here satisfied, by Proposition 2. The Z-method has terminated here at $x^o = 0$, a nonoptimal point. However, the MELP1 is applicable and, starting from the initial point $x^o = 0$, it gives the following results:

ℓ	x_1^ℓ	x_2^ℓ	x_3^ℓ	x_4^ℓ	x_5^ℓ	Active constraints	Pivot set Ω	Step size α_ℓ	$f^o(x^\ell)$
0	0	0	0	0	0	{1,2,4,5,6,7}	{2,6,7}	0.70711	9
1	0	0	0.70711	0.70711	0.70711	{1,3,4,5}	{3}	0.10138	3.42893
2	0	0	0.80849	0.69226	0.70711	{1,4,5}	ϕ	0.00739	3.41843
3	0	0	0.81587	0.69965	0.71449	{1,3,4,5}	{3}	0.02806	3.37735
4	0	0	0.84393	0.72513	0.68643	{1,4,5}	ϕ	0.00107	3.37513
5	0	0	0.84500	0.72619	0.68750	{1,3,4,5}	{3}	0.03226	3.36927
6	0	0	0.87726	0.69393	0.71726	{1,4,5}	ϕ	0.00142	3.36631
7	0	0	0.87869	0.69536	0.71868	{1,3,4,5}	{3}	0.02273	3.35859
8	0	0	0.90141	0.71670	0.69596	{1,4,5}	ϕ	0.00071	3.35710
9	0	0	0.90212	0.71741	0.69667	{1,3,4,5}	{3}	0.01888	3.35329
10	0	0	0.92100	0.69853	0.71462	{1,4,5}	ϕ	0.00049	3.35226
11	0	0	0.92149	0.69901	0.71511	{1,3,4,5}	{3}	0.01489	3.34968
12	0	0	0.93638	0.71332	0.70021	{1,4,5}	ϕ	0.00032	3.34903
13	0	0	0.93669	0.71364	0.70053	{1,3,4,5}	{3}	0.01201	3.34736
14	0	0	0.94871	0.70163	0.71216	{1,4,5}	ϕ	0.00020	3.34694
15	0	0	0.94890	0.70182	0.71236	{1,3,4,5}	{3}	0.00965	3.34590
16	0	0	0.95855	0.71123	0.70271	{1,4,5}	ϕ	0.00013	3.34562
17	0	0	0.95868	0.71135	0.70283	{1,3,4,5}	{3}	0.00779	3.34497
18	0	0	0.96646	0.70357	0.71046	{1,4,5}	ϕ	0.00010	3.34479
19	0	0	0.96655	0.70366	0.71055	{1,3,4,5}	{3}	0.00629	3.34432
20	0	0	0.97284	0.70984	0.70426	{1,4,5}	ϕ	0.00006	3.34420
21	0	0	0.97290	0.70989	0.70431	{1,3,4,5}	{3}	0.00509	3.34392
22	0	0	0.97798	0.70481	0.70933	{1,4,5}	ϕ	0.00004	3.34384
23	0	0	0.97803	0.70485	0.70937	{1,3,4,5}	{3}	0.00779	3.34363
24	0	0	0.98582	0.70485	0.70929	{1,4,5}	ϕ	0.00004	3.34357
25	0	0	0.98586	0.70489	0.70933				3.34335

By the "pivot set Ω" we have denoted, at every iteration x^ℓ, the biggest (by cardinality) subset Ω of $P(x^\ell)$ which generates the

direction of decrease $d(\Omega) \neq 0$ (as a solution of linear program (L,Ω)). The sequence x^{ℓ} converges to the optimal solution

$$x_1^* = 0, \quad x_2^* = 0, \quad x_3^* = 1, \quad x_4^* = \frac{\sqrt{2}}{2}, \quad x_5^* = \frac{\sqrt{2}}{2}$$

with the optimal value $f(x^*) = 9 - 4\sqrt{2}$.

The same problem will now be solved using the MELP2. Starting from the same initial approximation $x^o = 0$, the following results are obtained:

ℓ	x_1^{ℓ}	x_2^{ℓ}	x_3^{ℓ}	x_4^{ℓ}	x_5^{ℓ}	Active constraints	Pivot set Ω	Step size α_{ℓ}	$f^o(x^{\ell})$
0	0	0	0	0	0	{1,2,4,5,6,7}	{2,6,7}	0.70711	9
1	0	0	0.70711	0.70711	0.70711	{1,3,4,5}	ϕ	0.29289	3.42893
2	0	0	1.00000	0.70711	0.70711				3.34314

The optimal solution, correct to five decimal places, is reached in only two iterations! At the initial approximation $x^o = 0$, two noneliminated subsets of largest cardinality are $\Omega_1 = \{2,6,7\}$ and $\Omega_2 = \{2,6\}$. Since the corresponding optimal values of linear programs are here equal, i.e. $\lambda(\Omega_1) = \lambda(\Omega_2) = 1$, we choose Ω_1 to be the pivot set. (Whenever there is a tie, as in the above case, one can systematically choose the first Ω for the pivot.) At the next approximation x^1, the two noneliminated subsets are $\Omega_1 = \{3\}$ and $\Omega_2 = \phi$. Since $\lambda(\Omega_1) = 0.20711$ and

$\lambda(\Omega_2) = 0.58578$, we conclude that $\lambda(\Omega_2) > \lambda(\Omega_1)$ and choose Ω_2 to be the pivot set. Let us note that, at each iteration x^ℓ, there are here at most two noneliminated subsets Ω.

Example 9. The following program with 50 variables, 20 nonlinear convex and 5 linear constraints has been solved by the PFDM and compared with the Z_ε-method.

$$\text{Min } \frac{1}{2} \sum_{j=1}^{50} (x_j - \alpha_j)^2$$

s.t.

$$\exp(x_1 + x_2 + x_3) + 2\exp(x_1 - x_4 - x_5) \le 20.83$$

$$\exp(-x_6) + \exp(-x_6 + 2x_7 + x_8) + \exp(x_{10} - x_9) \le 2.05$$

$$\exp(-x_9 + 2x_{15}) + 2\exp(-x_{12}) \le 25$$

$$4\exp(5x_{11} + x_{17} - 2x_{18}) + \exp(x_{17}) \le 215.68$$

$$\exp(-x_{16}) + 3\exp(-x_{26}) + \exp(-x_{36}) \le 5$$

$$\exp(-x_{21}) + 7\exp(x_4 - x_{14}) + \exp(x_{22} + x_{24}) \le 142.7$$

$$\exp(x_5 + 0.1\, x_{19} - 2x_{28}) + 2\exp(-x_5 + x_{29}) \le 3$$

$$\exp(0.4(x_6 + x_{25}) + x_{27}) + 10\exp(-x_{23}) \le 10$$

$$10^{-2}\exp(x_{31} + x_{32} + x_{33}) + \exp(x_{34} - x_{35}) + \exp(-x_{31}) \le 5$$

$$10^{-3}\exp(0.5(x_{34} + x_{35}) + x_{36}) + \exp(x_{37} + x_{38}) \le 1.15$$

$$10^{-3}\exp(x_{39} + x_{40} + x_{41} + x_{42} + x_{43}) + \exp(-x_{44} - x_{45}) \le 0.29$$

$$\exp(x_{46} - 2x_{47} + x_{48}) + \exp(x_{49} - x_{50}) \le 2$$

$$\exp(x_{12} - x_{16}) + \exp(x_{12} + x_{16}) \le 10$$

$$x_6^4 - x_6 + \exp(x_{11}) + x_{11} \le -0.63$$

$$(x_{11} - 1)^2 + (x_{21} - 1)^2 + (x_{31} - 1)^2 \le 29$$

$$x_1^2 + x_2^2 + x_3^2 + 3x_{13} \le 8$$

$$\exp(-x_2 + x_5) + (x_{20} + x_{30})^2 \le 10^{-2}$$

$$(x_1 + 2x_4)^2 + \exp(-x_8) \le 10$$

$$(x_{16} + x_{17} + x_{18})^2 + 10^{-3} \exp(-x_{33}) \le 9$$

$$\exp(x_8) + (x_{48} - x_{49} + x_{50})^4 - x_8 \le 629$$

$$x_3 + 3x_6 + x_{12} - 7x_{20} - 2x_{30} \le 0$$

$$x_1 + x_7 - 3x_{29} - 4x_{41} - 4x_{42} \le -9$$

$$x_5 + x_{15} + x_{25} + x_{35} \le 9.5$$

$$x_{49} - x_{50} \le 0$$

$$4x_2 + 3x_{10} - 0.5x_{13} + x_{23} - 3x_{43} \le 7$$

where α_j, $j=1,\ldots 50$ are given below

-14	5	1	-9	-1	3	1	-1	5	0
-2	1	-33	0	0	-1	0	0	1	-7
0	0	1	0	0	-3	0	-2	-1	-2
0	0	0	0	0	-1	0	0	0	0
-4	-4	0	1	1	1	-2	1	2	-2

The initial feasible point x^o is

-3	5	0	1.5	-1	0	0	0	-2	-5
-1	0	-34	3	0.5	0	0	0	10	0
4	0	10	0	5	0	-10	0	-1	0
3	3	-6	5	5	0	-4	4	0	0
3	3	-3	6	6	5	5	5	5	5

and the optimal solution x^* is

-9	5	0	3	0	0	0	0	5	0
-2	0	-33	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	0

with the optimal value $f^o(x^*) = 149.005$. One can verify that x^* satisfies the Kuhn-Tucker condition with multipliers

$$\lambda_5 = \lambda_7 = \lambda_{12} = \lambda_{18} = \lambda_{21} = \lambda_{22} = \lambda_{24} = 1,\ \lambda = 0 \text{ otherwise .}$$

Value of objective function

Iteration	PFDM	z_ε
0	574.755	574.755
1	551.615	574.755
5	407.631	570.873
10	310.172	440.147
50	149.795	391.386
100		350.089
400		246.635
800		199.518

The MELP and the PFDM have been tested on the IBM 360/75 computers at the Technion and at McGill University.

REFERENCES

[1] A. Ben-Israel and A. Ben-Tal, "On a characterization of optimality in convex programming", Mathematical Programming 11 (1976), 81-88.

[2] A. Ben-Israel, A. Ben-Tal and A. Charnes, "New characterizations of Pareto optimality", *Econometrica* 45 (1977), 811-820.

[3] A. Ben-Tal, A. Ben-Israel and S. Zlobec, "Characterization of optimality in convex programming without a constraint qualification", *Journal of Optimization Theory and Application* 20 (1976), 417-437

[4] A. Charnes and W.W. Cooper, *Management Models in Industrial Applications of Linear Programming,* vol. I (J. Wiley, New York, 1961).

[5] A.R. Colville, "A comparative study on nonlinear programming codes", Rept. 320-2949, IBM New York Scince Center (1968).

[6] G. Debreu, *Theory of Value, an Axiomatic Analysis of Economic Equilibrium.* (J. Wiley, New York, 1959).

[7] V.V. Fedorov, "Optimization problems with an ordered set of constraints", Z. *Vic. Mat. i Mat. Fiz.* 15 (1975),1126-1137.

[8] D.M. Himmelblau, *Applied Nonlinear Programming.* (McGraw-Hill, New York, 1972).

[9] P. Johan, "Algorithms for the vector maximization problem", *Mathematical Programming,* 2 (1972), 207-229.

[10] S. Karlin, *Mathematical Methods and Theory in Games, Programming, and Economics,* vol. I. (Addison-Wesley, Reading, Mass., 1959).

[11] H.W. Kuhn and A.W. Tucker, "Nonlinear programming", in: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* Ed. J. Neyman (University of California, Berkeley, Calif. 1951).

[12] B.D. Lebedev, V.V. Podinovskii and R.S. Styrikovic, "An optimization problem with respect to the ordered totality of criteria", *Ekonomika i Matematiceskie Metody,* 7 (1971), 612-616.

[13] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming.* (Addison-Wesley, Reading, Mass., 1973).

[14] O. Mangasarian, *Nonlinear Programming.* (McGraw-Hill, New York, 1969).

[15] B.A. Murtagh and R.W.H. Sargent, "A constrained minimization method with quadratic convergence", in *Optimization*, Ed. R. Fletcher (Academic Press, London, 1969).

[16] M. Osborne, "Topics in optimization", Research Report STAN-CS-72-279 (Computer Science Department, Stanford University, Stanford, California, 1972).

[17] V.V. Podinovski and V.M. Gavrilov *Optimization with Respect to Successive Criteria, Siviet Radio*, Moscow, 1975. (In Russian.)

[18] M.J.D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives", *Computer Journal*, 7 (1964), 155-162.

[19] S.M. Robinson, "A quadratically convergent algorithm for general non-linear programming problems", *Mathematical Programming*, 3 (1972), 145-156.

[20] J. Rosen, "The gradient projection method for nonlinear programming, I. Linear constraints", *Journal of the Society for Industrial and Applied Mathematics*, 8 (1960), 181-217.

[21] J. Rosen, "The gradient projection method for nonlinear programming, II. Nonlinear constraints", *Journal of the Society for Industrial and Applied Mathematics*, 9 (1961), 514-532.

[22] D.M. Topkis and A.F. Veinott, "On the convergence of some feasible directions algorithms for nonlinear programming", *SIAM Journal on Control*, 5 (1967), 268-279.

[23] W.I. Zangwill, *Nonlinear Programming, A Unified Approach*. (Prentice Hall, Englewood Cliffs, N.J., 1969).

[24] G. Zoutendijk, *Methods of Feasible Directions, A Study in Linear and Non-linear Programming*. (Elsevier, New York, 1960).

[25] G. Zoutendijk, *Mathematical Programming Methods*. (North-Holland, Amsterdam, 1976).

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Center for Cybernetic Studies
The University of Texas

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
6 A New Class of Feasible Direction Methods

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
10 A./Ben-Tal
S./Zlobec

14 CCS-311

6. REPORT DATE
11 Sep[redacted]77

7a. TOTAL NO. OF PAGES
12 [redacted] 49 P.

7b. NO. OF REFS
25

8a. CONTRACT OR GRANT NO.
15 N00014-67-A-0126-0009
N00014-75-C-0569

b. PROJECT NO.
NR047-021

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
9 Center for Cybernetic Studies
Research Report CCS 311

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office of Naval Research (Code 434)
Washington, D.C.

13. ABSTRACT

This paper introduces a new class of feasible direction methods for solving differentiable convex programs with nonlinear convex constraints. Unlike many presently used methods, the ones introduced here are ~~not~~ based on the Fritz John or the Kuhn-Tucker theory but rather on two recent characterizations of optimality without a constraint qualification. The new methods are capable of generating feasible directions of descent along the boundary of the feasible set and they consistently give directions of steeper descent than many popular methods. This is achieved by solving only one linear program at each iteration. The new methods are particularly useful in solving large sparse convex programs; some of the programs tested had 100 variables and 50 nonlinear constraints. Moreover, the new methods are applicable whether or not Slater's condition or any other constraint qualification is satisfied.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

406197

Unclassified
Security Classification

hc

A-31408

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Feasible Direction Methods						
Convex Programming Algorithms						
Constraint qualification-free Method						
Boundary Direction Algorithms						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0102-014-6600

Unclassified
Security Classification

A-31409